WHAT IS CLAIMED IS:

1. An isolated ancestral viral gene sequence, and fragments thereof, wherein the sequence is a determined founder sequence of a highly diverse viral strain, subtype or group.

2. The sequence of claim 1, wherein the ancestral viral gene sequence is an HIV-1 ancestral viral gene sequence, an HIV-2 ancestral viral gene sequence, or a Hepatitis C ancestral viral gene sequence.

3. The sequence of claim 1, wherein the ancestral viral gene sequence is of HIV-1 subtype A, B, C, D, E, F, G, H, J, AG, or AGI; HIV-1 Group M, N, or O; or HIV-2 subtype A or B.

4. The sequence of claim 1, wherein the ancestral viral gene sequence is of widely dispersed HIV-1 variants, geographically-restricted HIV-1 variants, widely dispersed HIV-2 variants, or geographically-restricted HIV-2 variants.

5. The sequence of claim 1, wherein the ancestor viral gene sequence is an *env* gene or a *gag* gene.

6. The sequence of claim 1, wherein the ancestral viral gene sequence is more closely related, on average, to a gene sequence of any given circulating virus than to any other variant.

7. The sequence of claim 1, wherein the ancestor sequence has at least 70% identity with the sequence set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:6 or an mrca sequence set forth in Figures 9 to 17 or 27 to 35, and wherein the sequence does not have 100% identity with any circulating variant.

8. The sequence of claim 1, which encodes an ancestor protein of SEQ ID NO:2 or SEQ ID NO:4.

9. An isolated ancestor protein or antigenic fragment thereof from HIV-1, HIV-2 or Hepatitis C.

10. The sequence of claim 9, wherein the ancestor protein sequence has at least 70% identity with the sequence set forth in SEQ ID NO:2, SEQ ID NO:4 or an mrca sequence set forth in Figures 18 to 26 or 36 to 44, and wherein the sequence does not have 100% identity with any circulating variant.

11. The isolated ancestor protein of claim 9, which comprises the contiguous sequence of SEQ ID NO:2 or SEQ ID NO:4 or an mcra sequence set forth in one of Figures 18 to 26 or 36 to 44.

12. The isolated ancestor protein of claim 9, which is the ancestor protein of HIV-1 subtype A, B, C, D, E, F, G, H, J, AG, or AGI; Group M, N, or O; or HIV-2 subtype A or B.

13. The isolated ancestor protein of claim 12, which is at least 10 contiguous amino acids of an HIV-1 subtype B env ancestor protein or HIV-1 subtype C env ancestor protein.

14. The isolated ancestor protein of claim 9, which is gag or env protein.

15. An isolated expression construct comprising the following operably linked elements:

a transcriptional promoter;

a nucleic acid encoding an ancestor or COT protein; and

a transcriptional terminator.

16. The expression construct of claim 15, wherein the nucleic acid encodes SEQ ID NO:2 or SEQ ID NO:4.

17. The expression construct of claim 15, wherein the nucleic acid is the sequence set forth as SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:6.

18. The expression construct of claim 13, wherein the nucleic acid sequence is optimized for expression in a host cell.

19. The expression construct of claim 15, wherein the transcriptional promoter is a heterologous promoter.

20. The expression construct of claim 19, wherein the promoter is a cytomegalovirus promoter.

21. A cultured prokaryotic or eukaryotic cell transformed or transfected with the expression construct of claim 15.

22. The eukaryotic cell of claim 21, which is a mammalian cell.

23. The eukaryotic cell of claim 21, wherein the nucleic acid encodes the ancestor protein of SEQ ID NO: 2 or SEQ ID NO: 4.

24. The eukaryotic cell of claim 21, wherein the nucleic acid is the sequence set forth as SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5 or SEQ ID NO:6.

25. The prokaryotic cell of claim 21, which is an *E. coli* cell.

26. The eukaryotic cell of claim 21, which is an *S. cerevisiae* cell.

27. The eukaryotic cell of claim 21, which is a human cell.

28. A vector comprising the expression construct of claim 15.

29. The vector of claim 28, wherein the nucleic acid sequence is operably linked to a Semliki Forest Virus replicon, and wherein the resulting recombinant replicon is operably linked to a cytomegalovirus promoter.

30. An isolated host cell comprising the vector of claim 28.

31. A composition for inducing an immune response in a mammal comprising a highly diverse viral ancestor protein or an immunogenic fragment of an ancestor or COT protein.

32. The composition of claim 31, wherein the fragment is derived from the sequence set forth in SEQ ID NO:2 or SEQ ID NO:4.

33. The composition of claim 31, wherein the viral ancestor protein is from HIV-1 or HIV-2.

34. The composition of claim 31, which is a vaccine.

35. An isolated antibody that binds specifically to a viral ancestor protein and that binds specifically to a plurality of circulating descendant viral ancestor proteins.

36. The antibody of claim 35, wherein the ancestor protein is from HIV-1, HIV-2, or Hepatitis C.

37. The antibody of claim 35, which is a monoclonal antibody or antigen binding fragment thereof.

38. The isolated antibody of claim 35, wherein the antibody is a humanized monoclonal antibody.

39. The antibody of claim 35, wherein the antibody or antigen binding fragment thereof is a single chain antibody, a chimeric antibody, a single heavy chain antibody, an antigen binding $F(ab')_2$ fragment, an antigen binding Fab' fragment, an antigen binding Fab fragment, or an antigen binding Fv fragment.

40. A method of preparing an ancestral viral amino acid sequence, the method comprising:

(a) selecting circulating viral sequences of a highly diverse virus;

(b) determining an ancestral viral sequence by maximum likelihood phylogeny analysis that is a most recent common ancestor of the circulating viral sequences, the ancestral viral sequence representative of the evolutionary center of an evolutionary tree of the circulating viral sequences; and

(c) synthesizing a viral sequence that is not 100% identical to any of the circulating viral sequences but whose deduced amino acid sequence is at least 70% identical to any of them.

41. The method of claim 40, wherein the circulating viral sequences are from HIV-infected patients.

42. The method of claim 40, wherein the circulating viral sequences are from HIV-1, HIV-2 or Hepatitis C.

43. The method of claim 40, further comprising testing fragments in an assay for immunogenicity.

44. The method of claim 40, when the maximum likelihood phylogeny analysis includes coalescent likelihood analysis.

45. An isolated COT viral gene sequence, and fragments thereof, wherein the sequence is a determined sequence of a highly diverse viral strain, subtype or group.

46. The COT viral gene sequence of claim 45, wherein the COT viral gene sequence is an HIV-1 viral gene sequence, an HIV-2 viral gene sequence, or a Hepatitis C viral gene sequence.

47. The sequence of claim 45, wherein the COT viral gene sequence is of HIV-1 subtype A, B, C, D, E, F, G, H, J, AG, or AGI; HIV-1 Group M, N, or O; or HIV-2 subtype A or B.

48. The sequence of claim 45, wherein the COT viral sequence has at least 70% identity with an LScot or MMcot sequence set forth in Figures 9 to 17 or 27 to 35, and wherein the sequence does not have 100% identity with any circulating variant.

49. The sequence of claim 45, wherein the COT viral gene sequence is of widely dispersed HIV-1 variants, geographically-restricted HIV-1 variants, widely dispersed HIV-2 variants, or geographically-restricted HIV-2 variants.

50. The sequence of claim 45, wherein the COT viral gene sequence is an env gene or a gag gene.

51. The sequence of claim 45, which encodes a COT protein.

52. An isolated COT protein or immunogenic fragment thereof.

53. The isolated COT protein of claim 52, which is the COT protein of HIV-1 subtype A, B, C, D, E, F, G, H, J, AG, or AGI; Group M, N, or O; or HIV-2 subtype A or B.

54. The isolated COT protein of claim 52, which is at least 10 contiguous amino acids of an HIV-1 subtype B env COT protein or HIV-1 subtype C env COT protein.

55. The isolated COT protein of claim 52, wherein the COT protein sequence has at least 70% identity with an LScot or MMcot sequence set forth in Figures 18 to 26 or 36 to 44, and wherein the sequence does not have 100% identity with any circulating variant.

56. The isolated COT protein of claim 52, which is gag or env protein.

57. An isolated antibody that binds specifically to a COT viral protein and that binds specifically to a plurality of circulating COT viral proteins.

58. The antibody of claim 57, wherein the COT protein is from HIV-1, HIV-2, or Hepatitis C.

59. A method of preparing a COT viral amino acid sequence, the method comprising:

(a) selecting circulating viral sequences of a highly diverse virus;

(b) determining a COT viral sequence.

60. The method of claim 59, wherein the circulating viral sequences are from HIV-infected patients.

61. The method of claim 59, wherein the circulating viral sequences are from HIV-1, HIV-2 or Hepatitis C.

62. The method of claim 28, further comprising testing COT protein sequences or fragments thereof in an assay for immunogenicity.